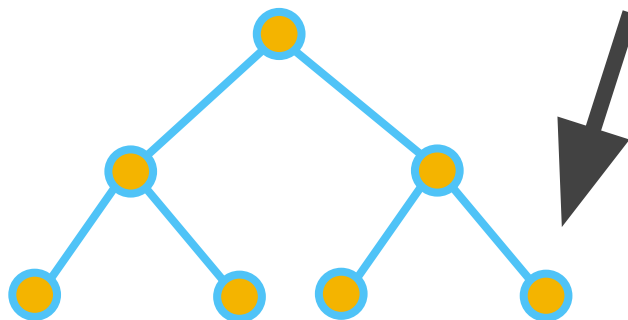
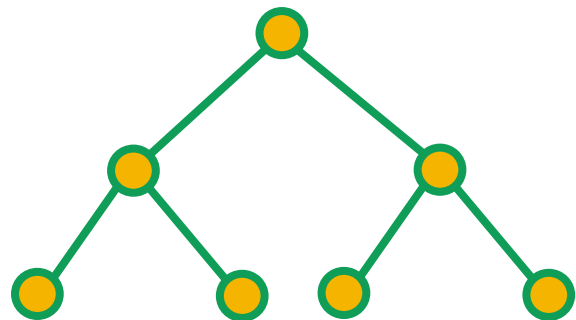
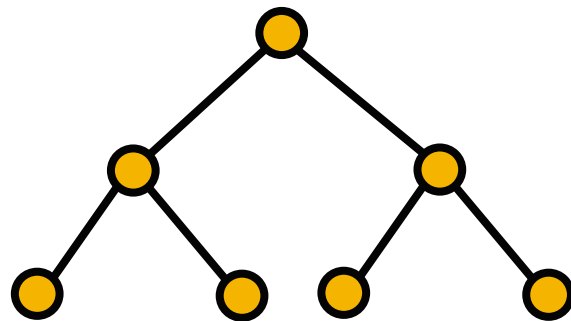
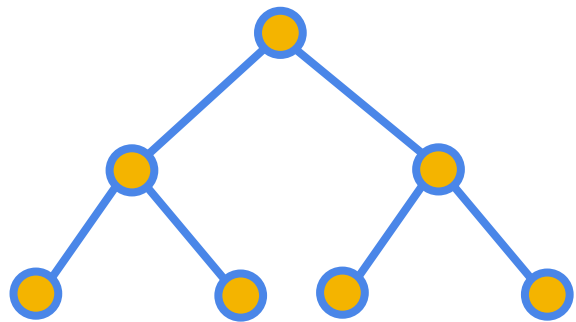
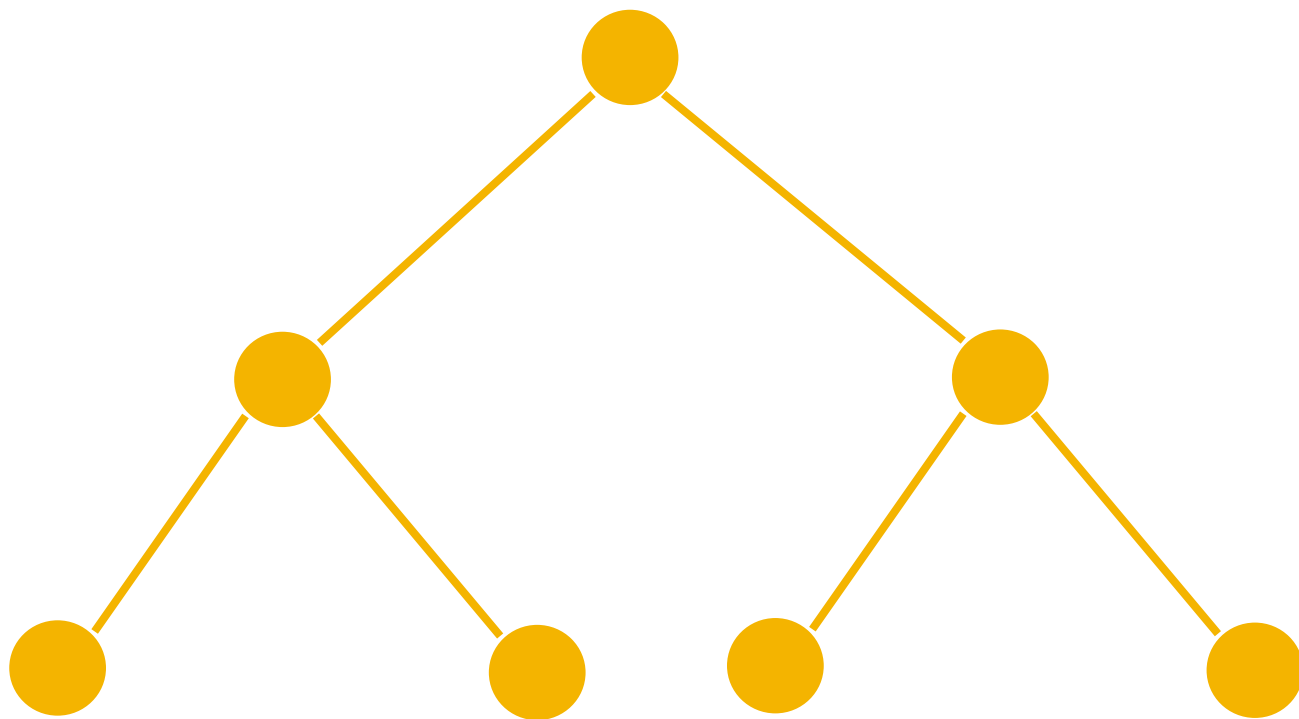
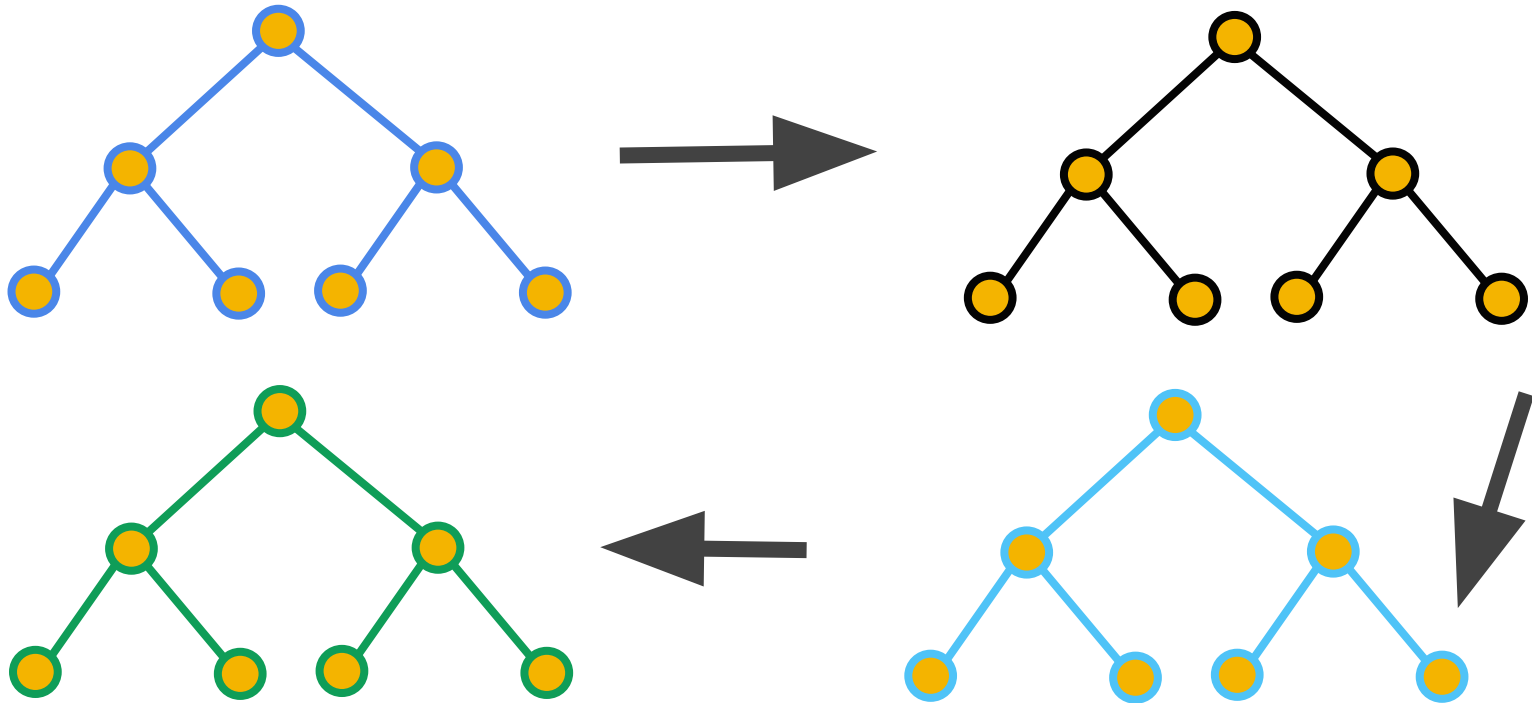


AdaBoost





Gradient Boosted Trees



Gradient boosting

Origins

- Can you improve a bad model?

Origins

- Can a weak learner can be modified to become better?

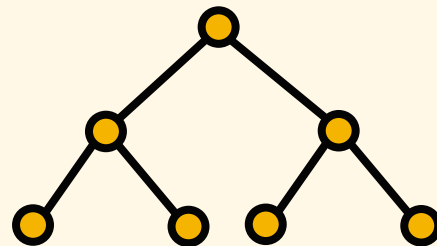
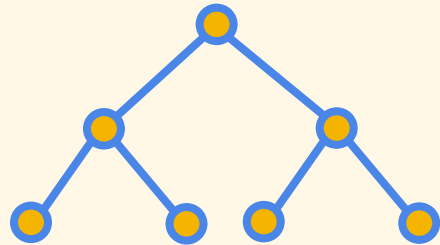
Weak learner

- Accuracy is at least **slightly better** than random chance.

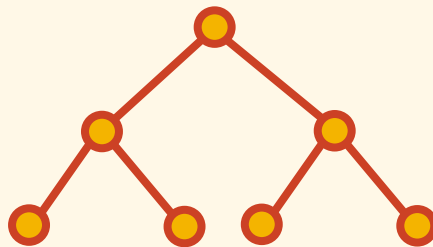
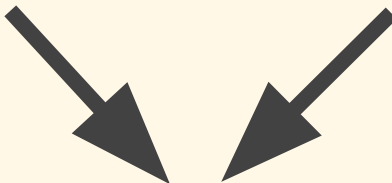
First Success

- AdaBoost

weak



weak

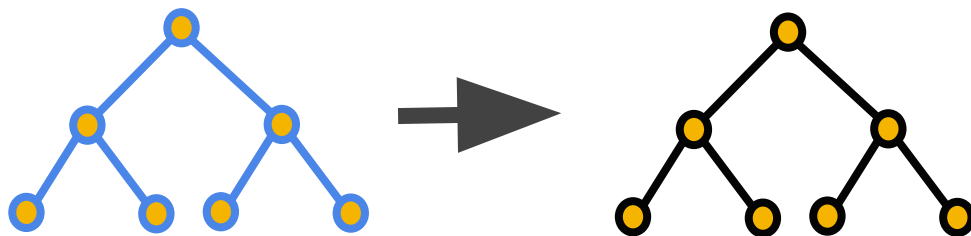


strong

AdaBoost

- weigh the observations
- put more weight on **hard to classify** instances
- less weight on others

AdaBoost



- New weak learners are added sequentially
- Focus training on the more difficult patterns

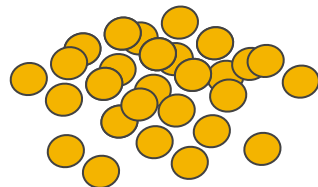
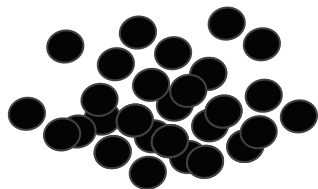
AdaBoost

- Samples that are hard to classify get increasing larger weights...
 - until the algorithm finds a model that **correctly classifies** these samples

AdaBoost

- Predictions are made by **majority vote** of the weak learners' predictions
 - weighted by their individual accuracy

AdaBoost



- Initially successful for **binary classification**

Gradient boosting

- Generalized AdaBoost
- Cast boosting as a **numerical optimization** problem

Goal

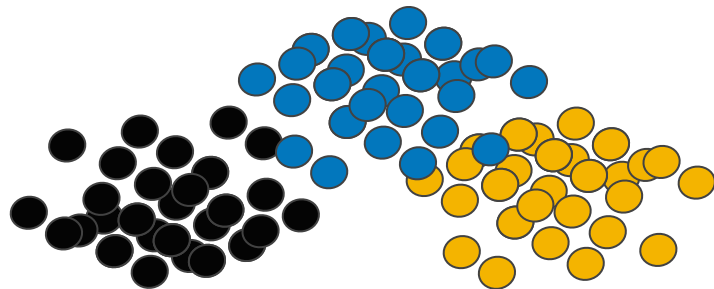
- reduce model **loss**

How?

- Add weak learners using a **gradient descent-like** process

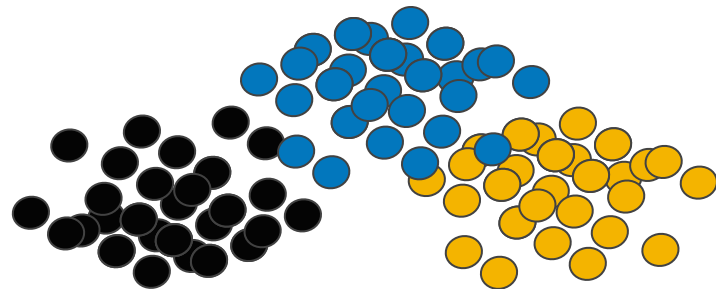
Stage-wise additive model

- 1 **new weak learner** added at a time
- existing weak learners in the model are **left unchanged.**



Stage-wise additive model

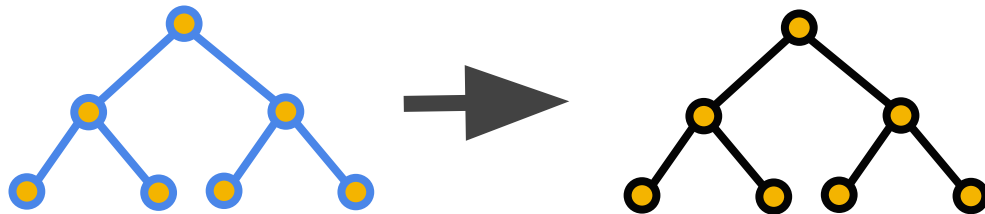
- support regression, multi-class classification and more.



Gradient boosting

- **ensemble** machine learning algorithms
- for classification or regression

Boosting



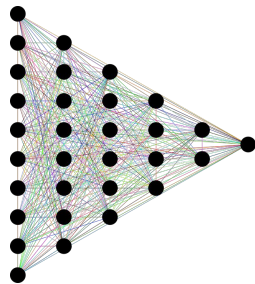
- **Trees** added **one at a time** to the ensemble
 - **fit to correct** the prediction errors made by prior models

Training

- Use any differentiable **loss function** and gradient descent **optimizer**

Gradient boosting

- loss gradient is **minimized** as the model is fit
 - like a neural network



Gradient boosting elements

- loss function
- weak learner
- additive model

Loss function

- To be optimized.
 - Mean squared error
 - ...

Weak learner

- To make predictions
- Decision trees

Additive model

- Add weak learners to **minimize** the loss function

AdaBoost ✓

